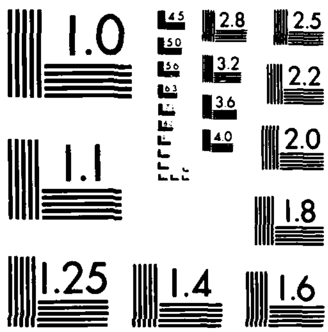
AD-A189 568 HOMOTOPIES: A PANACEA OR JUST ANOTHER METHOD?(U) 1/1
ROCHESTER UNIV NY DEPT OF COMPUTER SCIENCE I RIGOUTSOS
DEC 86 TR-201 DACA76-85-C-0001

UNCLASSIFIED F/G 12/9 NL



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

Homotopies: A Panacea or Just Another Method?

Isidore Rigoutsos
Department of Computer Science
The University of Rochester
Rochester, NY 14627

TR 201
December 1986

rochester

Department of Computer Science
University of Rochester
Rochester, New York 14627

(4)

Homotopies: A Panacea or Just Another Method?

Isidore Rigoutsos
Department of Computer Science
The University of Rochester
Rochester, NY 14627

TR 201
December 1986

Abstract

A general framework is presented for the solving of non-linear equations. Also, a discussion about its potential applications in the field of computer vision is made and illustrated by an example that shows how one can relate the solutions to the shape-from-shading problem through scale space. The methods presented seem to have been known since the 19th century; however, it was not until 1953 that the first practical applications of the relevant idea appeared.

This research was supported in part by the U.S. Army Engineering Topographic Laboratories Research Grant DACA76-85-C-0001. The Xerox Corporation University Grants Program provided the equipment used in the preparation of this paper.

A189 568

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER TR 201	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Homotopies: A Panacea or Just Another Method?		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Isidore Rigoutsos		8. CONTRACT OR GRANT NUMBER(s) DACA76-85-C-0001
9. PERFORMING ORGANIZATION NAME AND ADDRESS Computer Science Department University of Rochester Rochester, NY 14627		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Defense Advanced Research Project Agency 1400 Wilson Blvd. Arlington, VA 22209		12. REPORT DATE December 1986 13. NUMBER OF PAGES 24
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Office of Naval Research Information Systems Arlington, VA 22217		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited.

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution/
Availability Codes
Dist
A-1

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None

DTIC COPY INSPECTED

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

non-linear equations, shape-from-shading, scale space, homotopies, path following, Hilbert space, continuation methods

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A general framework is presented for the solving of non-linear equations. Also, a discussion about its potential applications in the field of computer vision is made and illustrated by an example that shows how one can relate the solutions to the shape-from-shading problem through scale space. The methods presented seem to have been known since the 19th century; however, it was not until 1953 that the first practical applications of the relevant idea appeared.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

HOMOTOPIES:
A panacea or just another method ?

Isidore Rigoutsos
Computer Science Department,
University of Rochester,
Rochester, NY 14627.

Abstract

A general framework is presented for the solving of non-linear equations. Also, a discussion about its potential applications in the field of Computer Vision is made and illustrated by an example that shows how one can relate the solutions to the Shape from Shading problem through Scale Space. The methods presented seem to have been known since the 19th century. However, it was not until 1953 that the first practical applications of the relevant idea appeared.

Index terms: Hilbert space, Homotopies, Path following, Scale Space, Shape from Shading.

1. Introduction

In a great many applications, one faces the problem of solving one or a system of non-linear equations. They could be non-linear equations determining the amount of life insurance that one should have ([16]), the pressure changes inside a gap between two objects moving very fast past to each other ([16]), the parameters involved in a calibration problem ([20]), the surface normal at a point (x, y) of an object in the shape from shading problem ([19]) etc.

The diversity of the above examples shows why solving non-linear equations is an important issue in current scientific applications, and explains the abundance of the numerical methods currently available for this purpose.

Most methods for solving one equation readily extend to methods that solve a system of linear equations. In general, these methods are characterized by (at least) two parts: an "iteration" part and a "convergence test" part. In what immediately follows, we are concerned about the first of these two parts.

Consider the equation $f(x) = 0$. The basic parameters of an iteration that solves the former, are:

-- a low-degree polynomial model for $f(x)$

-- an assumption about the behavior of $f(x)$

The most broadly used methods are:

a) *the bisection method*: it assumes no model for f(x) and tries to reduce the radius of the neighborhood around the solution of $f(x) = 0$ to the value 0, hence finding the solution x^* to the latter equation (Fig 1.a).

b) *Newton's method*: it assumes a linear model for f(x) and uses the slope of the line tangent at the point x of f(x) (Fig. 1.b).

c) *the secant method*: it is a variation of Newton's method that avoids computing the derivative of f(x); it also assumes a linear model based on the two most recent values of f(x) (Fig. 1.c).

d) *the regula falsi method*: it applies in the same context as the bisection method, except that now a straight line model is assumed (Fig 1.d).

e) *the fixed-point method*: one reformulates $f(x) = 0$ as $f(x) + x = x$ and finds the point of intersection of $g_1(x) = f(x) + x$ and $g_2(x) = x$, i.e. the fixed point of $g_1(x)$ (Fig. 1.e).

f) *the Muller method*: extension of the secant method that assumes a quadratic model (Fig. 1.f).

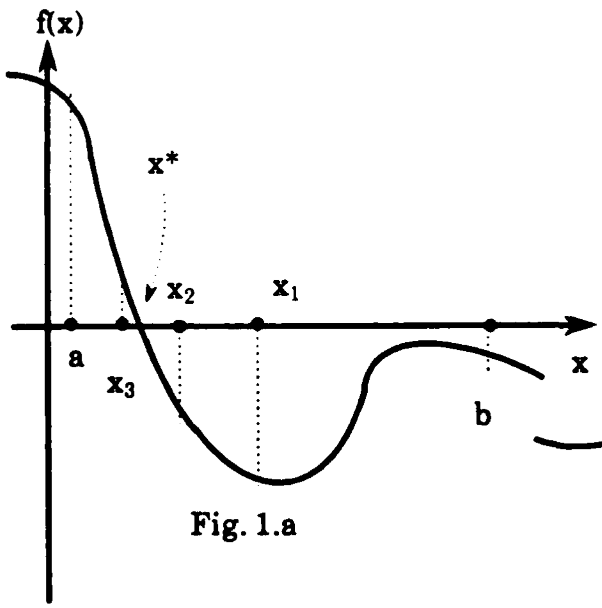


Fig. 1.a

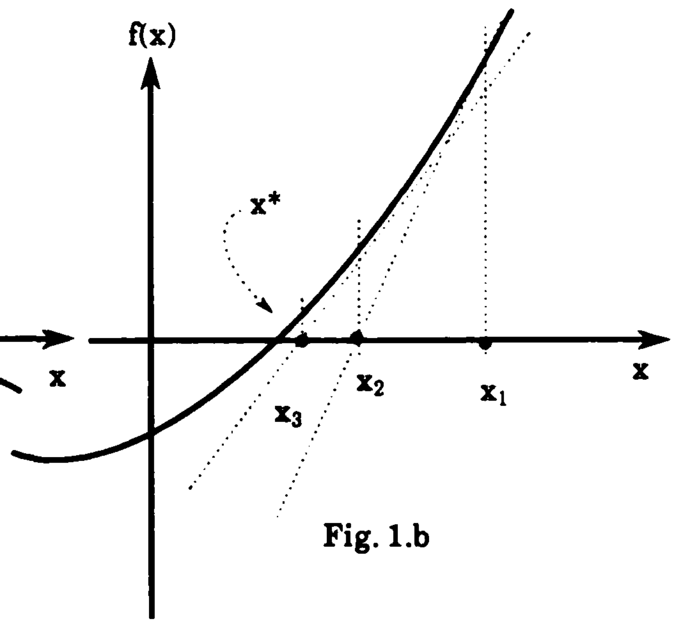


Fig. 1.b

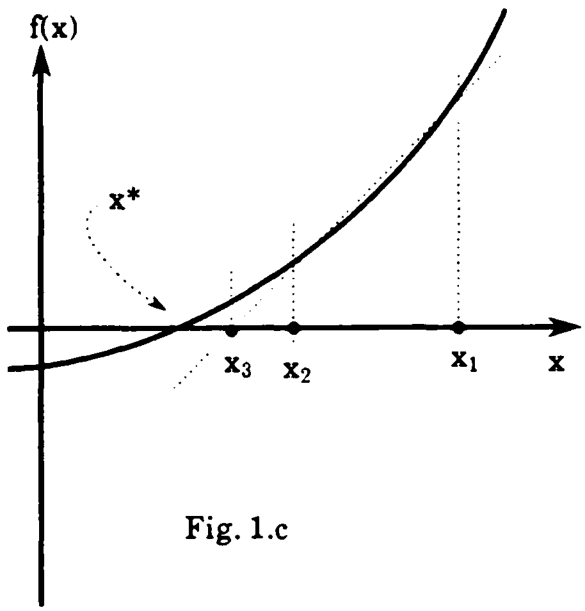


Fig. 1.c

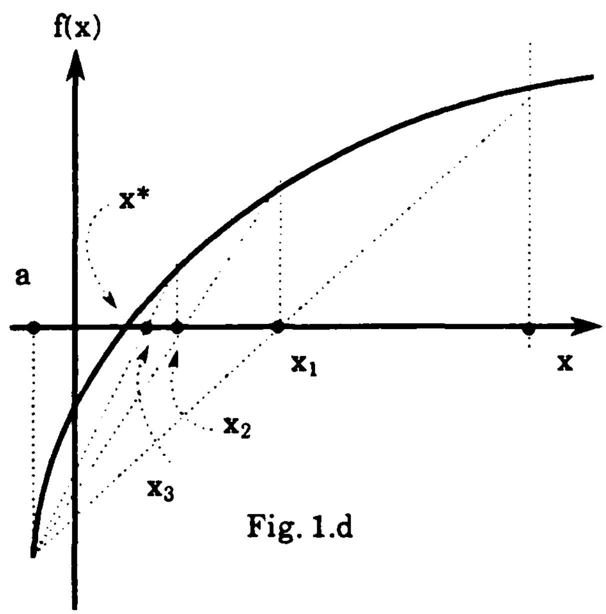


Fig. 1.d

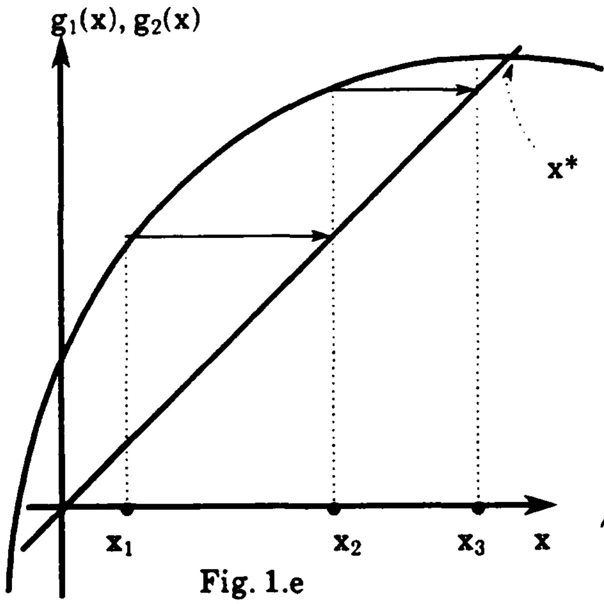


Fig. 1.e

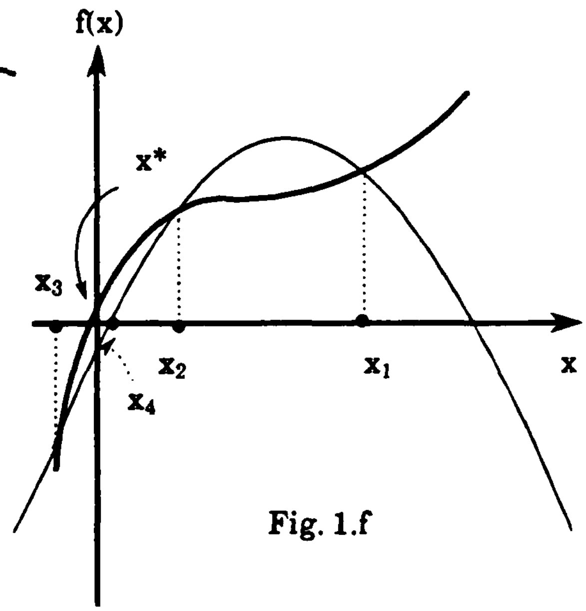


Fig. 1.f

2. Motivation and Previous work

The motivation for this study of the use of homotopies was a paper entitled "Signal Matching through Scale Space"([15]). This paper deals with the solution to the Signal Matching Problem through Scale Space and how it relates to the solution of some (relevant) known Computer Vision problems. The presentation was rather clear, however, the paper also contained a not-so-brief exposition of the authors' knowledge about solving problems of that kind. As a matter of fact, after having formulated the original problem as an optimization problem, they proceeded by suggesting the "stabilization methods" ([12]) as a candidate for the solving of this problem. This approach generated a non-linear system of equations; they then proposed the use of "continuation methods" ([11]) to solve this non-linear problem. However, they rejected this latter alternative and (apparently) used their intuition to provide us with a "*... a more attractive alternative ...*", that of gradient descent through scale space, which generates a set of equations "describing" the problem. However, once more time they rejected the equations that resulted as involving "*...high-dimensional spaces ...*", and suggested a new (approximately) equivalent to the former, new system of equations, which they finally solved.

Although intuition is always attractive, it would clearly be useful to provide a general framework in which the Signal Matching problem as well as all other optimization problems could be solved. Such a framework is presented in what follows. We start with a discussion about the use of problem imbedding during the last few years.

For about a century, scientists have known about imbedding methods. However, the homotopies did not specify how one can accurately calculate the solution of an equation or system of equations, or the fixed point of a function. Using them it was only possible to verify whether it is theoretically possible to reach a desired point. The first major step in the 20th century was taken by Davidenko ([18]). Davidenko was the first one that built on the homotopy idea by adding path following aspects to it. He provided a set of differential equations that "ruled" the path leading from an initial position to the desired point; these differential equations were applied to a great many deal of problems ranging from determinant evaluations and polynomial root finding to eigenvalue and boundary value problems ([13], [18], [9]). In general, his ideas led to algorithms that produced **differentiable** paths.

(*Note: By "path" we mean a piecewise differentiable curve in n-dimensions*)

Another more recent approach to path following was initiated by Scarf ([2]); Scarf did not use the homotopy approach but a new method based upon the notion of the so called "primitive sets". This approach led to algorithms generating **piecewise** linear paths. The underlying principle of both approaches is the same:

" ... *starting from* **here**, *follow a path that leads* **there...** ".

It has to be pointed out that there is no best path following approach. The approach that has to be taken by somebody, is closely related to the problem under consideration.

The layout of this paper is the following:
Section 3 is a tutorial on homotopies. **Section 4** discusses computational considerations of the methods that solve systems of differential equations. **Section 5** contains a discussion on how to get by local extrema, and finally **Section 6** illustrates the methods presented in **Section 3**, with an example showing how one can relate the solutions to the Shape from Shading problem through Scale Space, and also mentions other potential applications of the homotopy methods in the field of Computer Vision.

3. HOMOTOPIES (imbeddings)

3.1 A description

Let E be a real Hilbert space, and let $F: E \rightarrow E$ be a nonlinear operator.
Suppose that we are concerned with the numerical solution of the equation (actually the system of equations):

$$F(x) = 0 \qquad (1)$$

Suppose also that x^* is a solution of (1). A well-known method for approximating x^* is Newton's method. It consists of the calculation of a sequence of approximations $\{ x^k \}$, where

$$x^{k+1} = x^k - [F'(x^k)]^{-1} F(x^k), k = 0,1,2,3... \quad (2)$$

In (2), x^0 is a given approximation to x^*. The " ' " denotes the locally Lipschitz continuous, Frechet derivative of F at x.

Definition 1:

A mapping F*:* $D \subset E \rightarrow E$ *is Lipschitz continuous in* $D_0 \subset D$ *if there exists a constant c such that:*

$$\forall x, y \in D: \| F(y) - F(x) \| \leq c \| y - x \|$$

Definition 2:

Consider the mapping $H: E \rightarrow E$. H *is said to be Frechet differentiable at* $x \in$ **interior**(E) *if there is a linear operator* $A: E \rightarrow D \subset E$ *such that*

$$\lim_{h \rightarrow 0} \frac{\| H(x+h) - H(x) - A(h) \|}{\| h \|} = 0$$

The linear operator A *is denoted by* $F'(x)$ *and is called the* *Frechet derivative* *of* H *at* x.

Note: This last condition is a *uniformity* condition which allows most of the usual properties of derivatives in one dimension to be carried over to n dimensions. For example, if $H:E \rightarrow E$ is Frechet-differentiable at x, then H is continuous at x ([4]).

Unfortunately, Newton's method suffers from a problem; in particular, when x^0 is remote from x^* then the sequence $\{x^k\}$ defined by (2) will generally not converge to x^*. In such cases, homotopies (imbeddings) appear to be a useful tool for generating a sequence $\{x^k\}$ that converges to x^*.

Definition 3:

A homotopy is a mapping $H: E \times [t^0, t^1] \rightarrow E$, *for which* $H(x,t)$ *depends continuously on t, such that:*

$$H(x, t^0) = G(x) \qquad (3a)$$

$$H(x, t^1) = F(x) \qquad (3b)$$

with $H(x, t^0) = G(x) = 0$ *being an easily solvable system and* $G: E \rightarrow E$ *an (in general) non-linear operator.* $[t^0, t^1]$ *is a closed interval in* R.

F is imbedded (mapped) in the family of operators,

$$\{ H(\cdot, t) \ / t \in [t^0, t^1] \} \quad (4)$$

Now instead of the single problem that we initially had, we have the whole family of problems:

$$H(x, t) = \mathbf{0} \ / t \in [t^0, t^1] \quad (5)$$

Let $u^0 \in E$ be the solution of the equation $H(x, t^0) = \mathbf{0}$. Suppose that (5) has a unique solution $x = U(t)$ depending continuously on t. We therefore have

$$H(U(t), t) = \mathbf{0} \ / t \in [t^0, t^1] \quad (6)$$

and

$$U(t^0) = u^0, U(t^1) = x^*$$

Thus U defines a curve (path) in E with starting point u^0 and ending point equal to the solution x^* of (1).

3.2 An example

Let us now illustrate the homotopy method with an example. Consider the following nonlinear system

$$\begin{aligned} 2x_1^3 + 12x_1^2 + 8x_1 + 9x_2 + 36 &= 0 \\ 2x_1^2 + 3x_2 + 4 &= 0 \end{aligned}$$

Assume now that our initial system of equations is the following

$$\begin{aligned} 2x_1^3 + 8x_1 + 9x_2 &= 0 \\ 3x_2 &= 0 \end{aligned}$$

The unique (real) solution of this system is $(x_1, x_2) = (0, 0)$.
Now consider the real parameter $t \in [t^0, t^1] \equiv [0,1]$, and create a new system of equations such that this new system, at $t = 1$, becomes "equal" to the one we initially had. This new system is:

$$\begin{aligned} 2x_1^3 + 8x_1 + 9x_2 + t(12x_1^2 + 36) &= 0 \\ 3x_2 + t(2x_1^2 + 4) &= 0 \end{aligned}$$

Denoting the solution by $(x_1(t), x_2(t))$, we can see that $(x_1(0), x_2(0)) = (0,0)$ whereas $(x_1(1), x_2(1))$ is the "point" we wish to find. Elimination of x_2 from the above system results in

$$\begin{aligned} (x_1 + 3t)(2x_1^2 + 8) &= 0 \Rightarrow \\ x_1(t) &= -3t \\ x_2(t) &= -t(18t^2 + 4)/3 \end{aligned}$$

A graph (Fig. 2) of the solution(s) of the parameterized system, follows.
At $t=1$, $x_1 = -3$ and $x_2 = -22/3$, which are the non-zero solutions of the initial system. Although we could equally well solve the above non-linear system of

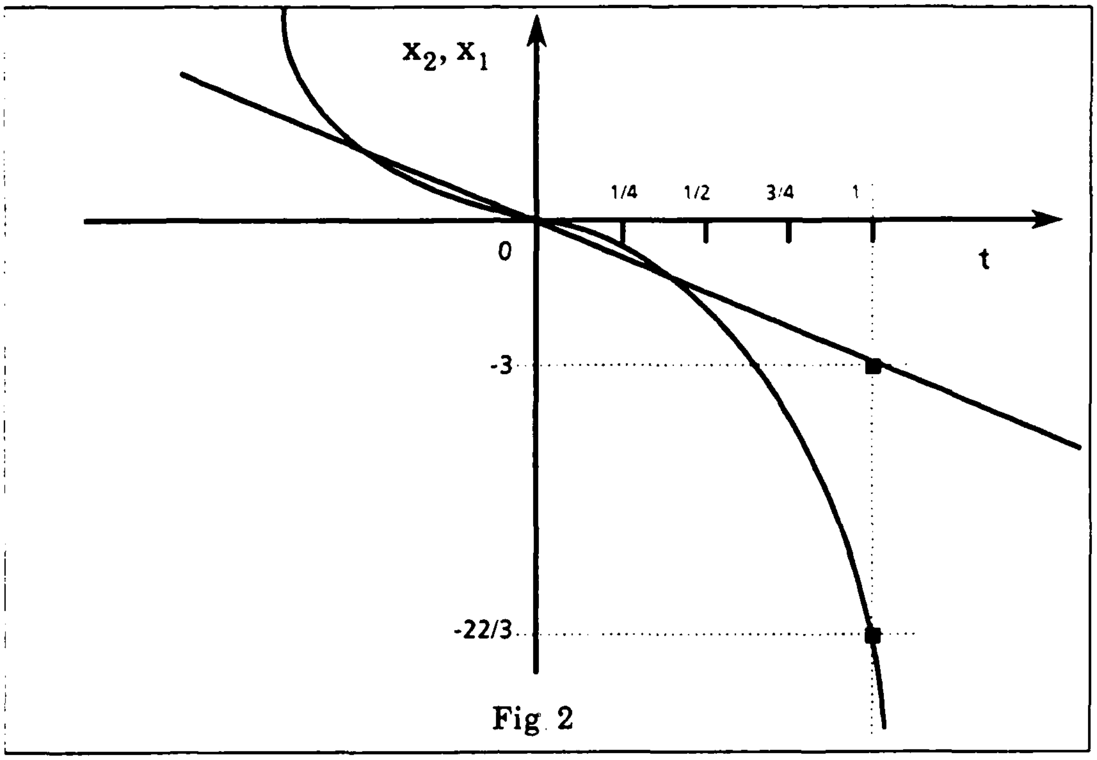


Fig. 2

equations without introducing the parameter t, the utility of the approach extends to *numerical* and not just *analytic* methods.

The simplicity of the previous example may make one think that "inspection" suffices and path following is superfluous. However, in n-dimensional spaces, where visual interpretations can hardly exist, we are forced to "inch" along the path.

Actually the exact path that occurs depends directly upon the selected homotopy function **H**(**x**,t).

We have specified the types of **G**, **H** in (3) but we have not elaborated upon the form of **H**(**x**,t). Thre are three common homotopies:

1) **Newton homotopy:** the form of this homotopy is:

$$\mathbf{H}(\mathbf{x},t) = \mathbf{F}(\mathbf{x}) - (t^1 - t)\mathbf{F}(\mathbf{x}^0), \text{ with } t^0 = 0 \text{ and } t^1 = 1$$

named after Sir Isaac Newton. Observe that $\mathbf{G}(\mathbf{x}) = \mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{x}^0)$ which has the obvious solution $\mathbf{x} = \mathbf{x}^0$.

2) **Fixed point homotopy**: the form of this homotopy is:

$$\mathbf{H}(\mathbf{x},t) = (t^1 - t)(\mathbf{x} - \mathbf{x}^0) + t\mathbf{F}(\mathbf{x}), \text{ with } t^0 = 0 \text{ and } t^1 = 1$$

A common characteristic of the first two homotopies is that they can be started at any point x^0.

3) **Linear homotopy**: the form of this third homotopy is:

$$H(x,t) = G(x) + t[F(x) - G(x)], \text{ with } t^0 = 0 \text{ and } t^1 = 1$$

i.e. this homotopy is a linear combination of $G(x)$ and $F(x)$. One can easily observe that the linear homotopy subsumes both the *Newton* (in which case $G(x) = x - x^0$) and the *Fixed Point* (in which case $G(x) = F(x) - F(x^0)$) homotopies; it is used whenever one wants the starting function $G(x)$ to have some special properties (as in the previous example).

3.3 Types of imbeddings

In the sequel we shall describe some ways in which the homotopy idea can be used in a numerical approximation of x^*, the latter being the solution of (1).

1) ***Discrete Imbedding****: this consists of successive numerical approximations of the* solutions of

$$H(x, t_i) = 0, \text{ for } i = 0, 1, 2, 3, \ldots N$$

N is an integer and the set $\{t_k\}$ is a partition of $[t^0, t^1]$. As a starting point for approximating the solution at $t = t_m$ the solution at $t = t_{m-1}$ is used. This method is also known under the name **continuation method** ([11], [15]). In this approach, if the distance $| t_k - t_{k-1} |$ is sufficiently small, then a sequence of solutions that converges to x^* can *possibly* be found.

2) **Transformation to an initial value problem**: this is a totally different approach to approximate x^*. Assume that the mapping $U(t)$ is continuously differentiable on $[t^0, t^1]$ and that H has locally Lipschitz continuous partial Frechet derivatives; then we obtain by differentiation of (6) with respect to t, a system of differential equations which together with the initial condition $U(t^0) = u^0$ form an initial value problem. The numerical solution of the latter can subsequently be obtained by application of a standard numerical integration procedure. The resulting solution can then be used as the starting point of an iterative method that approximates x^* more closely. This approach is attributed to Davidenko ([13], [18]) and is explained in the sequel in more detail.

3) **Iterative imbedding**: this is the least frequently used of all imbeddings. In this case one elaborates on the initial value problem idea. In particular, the homotopy is supposed to be a linear one (see above); then differentiation with respect to t leads to an initial value problem of the form:

$dx / dt = - [(t^1 - t)*\theta_1 G(x(t),x^0) + t*F'(x(t))]^{-1}*[-G(x(t),x^0) + F(x(t))]$

with $t^1 = 1$ and $t^0 = 0$ and $x(t^0) = x^0$ ([3]). θ_1 is the partial Frechet derivative of H with respect to t. The above system is solved by the following iterative process:

$$x^{k+1} = V(x^k),$$

where $x^1 \equiv V(x^0)$ is the solution of that system at $t=1$, by means of a numerical integration procedure. Then, x^0 is replaced by x^1, which in turn is replaced by the new derived solution x^2, and so on...

Up to this point we have ignored the following extremely important questions: a)existence, b)continuity, and c)uniquenes of the mentioned path.

3.4 Path Existence

Let us give some definitions first. Define

$$H^{-1} = \{ (x,t) / H(x,t) = 0 \} \quad (7)$$

to be the set of all solutions $(x,t) \in E \times [t^0,t^1]$ of the initial system of equations

$$H(x,t) = 0$$

In general, H^{-1} can be rather unconstrained. The points (x,t) that satisfy $H(x,t) = 0$ could be all over the place, experiencing no particular configuration. Let also

$$H^{-1}(t) = \{ x / H(x) = 0 \} \quad (8)$$

i.e. $H^{-1}(t)$ is the set of the solutions for a fixed value t. One can immediately observe that

$$H^{-1}(t^0) = \{ (x^0) / H(x,t^0) = G(x) = 0 \}$$

$$\text{and} \quad H^{-1}(t^1) = \{ x^* = x(t^1) / H(x,t^1) = F(x) = 0 \}$$

Therefore, we can see that the task of determining whether there are any paths of solutions becomes equivalent to determining whether H^{-1} is non-empty. One should keep in mind that there may be cases where H^{-1} consists of curves that may not be paths.We now proceed towards their elimination in order to ensure that H^{-1} contains only paths.

Recalling the definition of a Jacobian, we can write:

$$H' = (H'_x, \theta H / \theta t), \quad (9)$$

where

$$H'_x = \begin{bmatrix} \theta H_1/\theta x_1 & \theta H_1/\theta x_2 & \theta H_1/\theta x_3 & \ldots\ldots\theta H_1/\theta x_m \\ \theta H_2/\theta x_1 & \theta H_2/\theta x_2 & \theta H_3/\theta x_3 & \ldots\ldots\theta H_2/\theta x_m \\ \theta H_3/\theta x_1 & \theta H_3/\theta x_2 & \theta H_3/\theta x_3 & \ldots\ldots\theta H_3/\theta x_m \\ \vdots & \vdots & \vdots \quad \cdots & \vdots \\ \theta H_m/\theta x_1 & \theta H_m/\theta x_2 & \theta H_3/\theta x_3 & \ldots\ldots\theta H_m/\theta x_m \end{bmatrix} \qquad (10)$$

and

$$H'_t = (\theta H_1/\theta t \quad \theta H_2/\theta t \quad \theta H_3/\theta t \ \ldots\ldots\theta H_m/\theta t)^{\tau} \qquad (11)$$

At this point, we need to recall the well-known Implicit Function Theorem from the theory of Functional Analysis ([1])

Theorem: Let H: $E \rightarrow E$ be defined and continuously differentiable in a neighbourhood $E \times T$ of a point $(x_1, x_2, x_3, \ldots x_m, t) = (x_p, t_p)$ of $E \times [t^0, t^1]$ such that $H(x_p, t_p) = 0$ and the Jacobian H'_x is invertible at (x_p, t). Then there is an open neighborhood $W_0 \subset E$ of $(x_1, x_2, x_3, \ldots x_m)$ such that for any connected open neighborhood $W \subset W_0$ of $(x_1, x_2, x_3, \ldots x_m)$ there is a *unique* function g:

$$g: E \times [t^0, t^1] \text{ with } g(x_p) = t_p$$

defined, continuous, and continuously differentiable in W such that

$$H(x, g(x)) = 0 \quad \forall x \in W$$

For a proof of this theorem see [1], p. 265-268.

This theorem immediately excludes forks, crossings, splittings or other anomalies that may occur. We are therefore guaranteed that our path is extremely well behaved. One can proceed further and obtain the following corollary ([2]):

Corollary: Let H: $E \rightarrow E$ be defined and continuously differentiable in a neighbourhood $E \times T$ of a point $(x_1, x_2, x_3, \ldots x_m, t) = (x_p, t_p)$ of $E \times [t^0, t^1]$ such that $H(x_p, t_p) = 0$. Suppose now that for all such points **p** of $E \times [t^0, t^1]$ with $\mathbf{p} \in H^{-1}$, the

Jacobian H'_x is invertible at $p \equiv (x_p, t_p)$. Then H^{-1} consists only of continuous and continuously differentiable paths.

Hence the path existence and continuity can be ensured given the validity of the above theorem/corollary hypotheses. As for the uniqueness one can easily see that if the equation (1) has a unique solution, it follows from the implicit function theorem that the cardinality of H^{-1} equals 1. A more elaborate discussion on existence issues can also be found in [6].

3.5 Moving along paths

So far we dealt with results refering to the existence, uniqueness and continuity of the paths. But we have not discussed at all how one can find these paths.

3.5.1 *Davidenko's Approach*

Davidenko ([13], [18]) was the first one that introduced the differential equation as a means for solving $H(x,t) = 0$ with $t \in [t^0, t^1]$. His approach consisted of the following: he differentiated (1) with respect to parameter t and obtained the system of equations:

$$\sum_{i=1}^{m+1} \frac{\theta H}{\theta x_i} \frac{dx_i}{dt} + \frac{\theta H}{\theta t} = 0, \Leftrightarrow$$

$$\frac{dx}{dt} = -\frac{\theta H}{\theta t} \left[H'_x \right]^{-1}$$

Obviously, this system of equations requires that H'_x is non-singular. Furthermore, these equations used t as a parameter([4]). The use of t as a parameter is not a good idea, since it creates the following problem: "at a branching point, the matrix H'_x becomes singular and any integration method obviously fails" (see also the discussion about local extrema, in Section 5). One may want to adapt their method in such a way so that it is able to handle branching points (and local extrema with appropriate extensions). The first such adaptation was by Klopfenstein ([5]) who introduced a natural parameterization using the arc length of the solution locus; the latter parameterization is apparently a powerful technique and does not suffer from the above mentioned problems.

3.5.2 *The Basic differential equations*

Let us now use the parameter 's' to denote the distance moved along the path (length of the arc). Let therefore

$$Y(s) \equiv (y_1(s), y_2(s), y_3(s), \ldots y_m(s), y_{m+1}(s)) = (x(s),t(s)) \qquad (12)$$

where we denote x_i by y_i, $\forall\, i \in \{1, 2, 3, \ldots m\}$, and t by y_{m+1}, for clarity. Observe that the point Y(s) tells us where we are, after having travelled a distance s along the path. Since as s varies Y = Y(s) describes a path in H^{-1}, we conclude that

$$H(Y(s)) = 0 \qquad (13)$$

Differentiation of the latter with respect to s, and application of the chain rule, gives

$$\sum_{i=1}^{m+1} \frac{\theta H}{\theta y_i(s)} \frac{dy_i(s)}{ds} = 0 \qquad (14)$$

where $\theta H / \theta y_i(s)$ is the i-th column of the Jacobian H'. Usually, H is an m-dimensional vector, $H=(H_1, H_2, \ldots, H_m)$, in which case this latter equation is a system of m linear quations in the m+1 unknowns dy_i/ds, $i=1,2,3, \ldots m+1$. The (m+1)th equation is the following:

$$\left(\frac{dy_1}{ds}\right)^2 + \left(\frac{dy_2}{ds}\right)^2 + \left(\frac{dy_3}{ds}\right)^2 + \ldots + \left(\frac{dy_{m+1}}{ds}\right)^2 = 1 \qquad (15)$$

which results from the very definition of the length of the arc ([17]).

In what follows, we adopt the following notation:

<u>*Notation*</u>: *By* H $\in C^k$ *we mean that* H *has continous derivatives up to the k-th order.*

<u>*Definition 4:*</u> *An operator* H *is called regular if the Jacobian matrix* H'*(*x*,t) has a rank equal to m (i.e. equal to the size of the shortest side of the Jacobian matrix).*

The Basic Differential Equations theorem: Let H: E×E, H $\in C^2$, and H be regular. Given a starting point y^0 in H^{-1}, the solution of the "basic differential equations"

$$\frac{dy_i}{ds} = (-1)^i * \det\left[{}^iH'(y)\right], \; i=1,2,3,4 \ldots m+1$$

starting from $y(s) = y^0$ is unique and determines a path in H^{-1} ([2]).

(Note: by ${}^{i}H'(y)$ we denote the m x m matrix that results if we delete the i-th column of the Jacobian matrix $H'(y)$.)

The Basic Differential Equations provide a general means to follow a path in H^{-1} and therefore obtain a solution point $x^* \in H^{-1}(t^1)$. However, the Basic Differential Equations approach is far more than merely a means to solve a system of differential equations. They also provide immediate and important information concerning the direction (≡ orientation) of the path.

3.6 Which way are we travelling – coming or going?

While dealing the system of the Basic Differential Equations it is important to consider certain aspects of the problem. One may naively think that with increasing s, t increases as well. Unfortunately, this is not the case as it can be easily seen from the following figure (Fig. 3) reproduced from [10]. Suppose that we are at $t = t^0$ and desire to increase t. If $t > 0$ then increasing s will increase t. Respectively, if $t < 0$ the opposite is true. Hence, we must first check the sign of t. Similar phenomena would occur if we moved from some position (x^2,t^2) to a new one (x^3,t^3) along the path; which way to change s is not clear beforehand.

3.7 Using which approach should one try to "move" ?

The system of equations that results from Davidenko's approach can be solved whenever H'_x has a rank equal to m (see above). On the other hand, if H is regular we may write $y=(x,t)$ as a function of s and obtain the set of Basic Differential Equations.

The assumption of the "regularity of H'_x" is a rather mild one and holds almost always. No matter which system we decide to solve, both systems must (and will) generate the same path since they were derived from the same homotopy.

4. Computational considerations

Davidenko's approach can be used whenever there is a problem during the evaluation of the determinants involved in the Basic Differential Equation approach. However, one can also face some problems using the former method since the evaluation of the inverse Jacobian H'_x may involve too much computational effort. The usual tricks that are used in this case are:

1) use difference formulas instead of calculating the derivatives of H'_x.

2) do not evaluate the Jacobian at every step but rather every few steps.

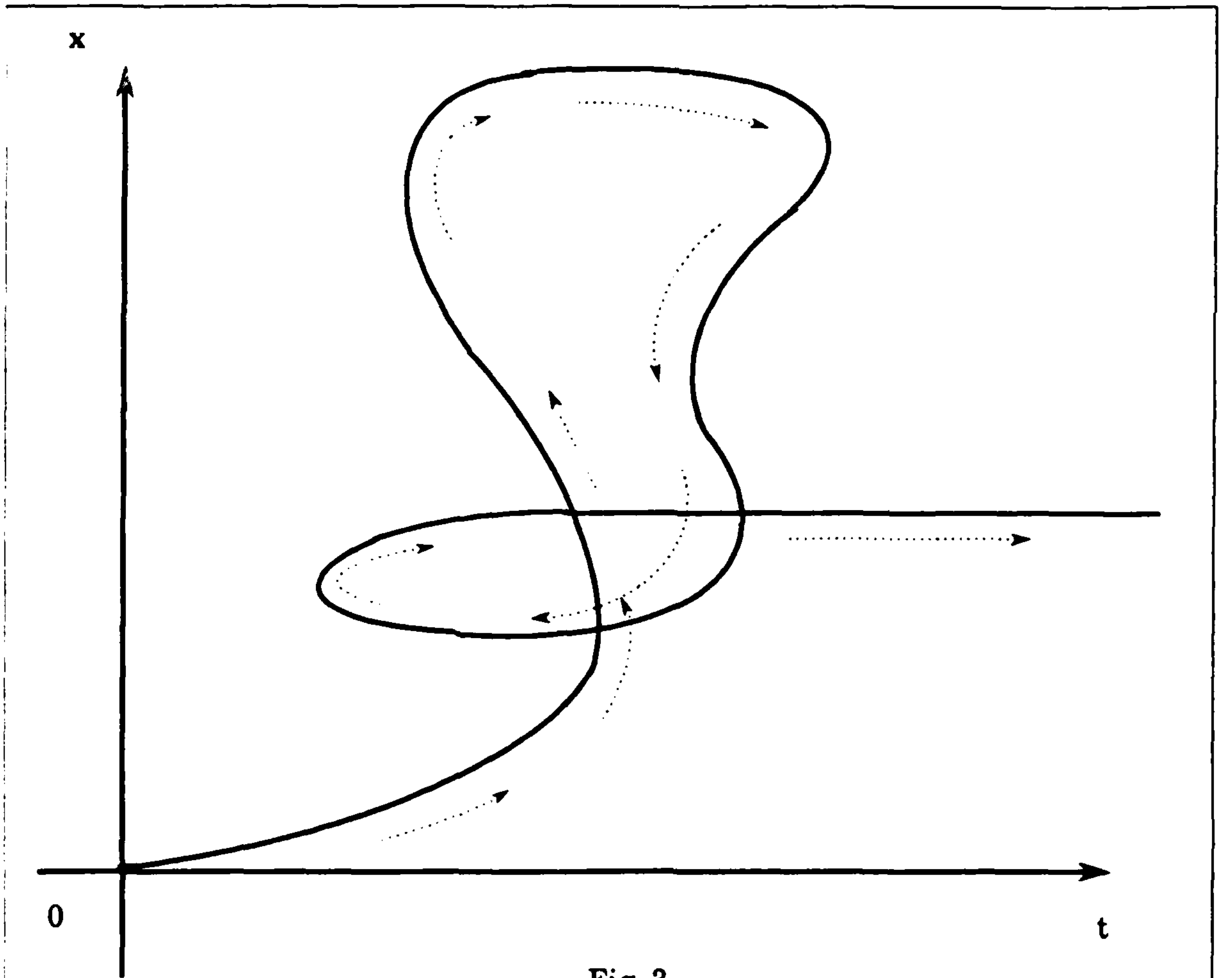


Fig. 3

4.1 Choosing the correct numerical method

Recall the Euler method formula for the general initial value problem:

$$\frac{dy}{dx} = f(x,y), \text{ where } y \in R^k,\ x \in R \text{ and } y(0) \text{ is known}$$

Euler's formula is

$$y^{k+1} = y^k + f(x^k, y^k)\,(y^{k+1} - y^k), \text{ where } 0 < y1 < y2 < y3 < \ldots < yk < yk+1 \wedge y(0) = y^0$$

Euler's method is a simple method based on the linear model and has a characteristic low accuracy. In practice the round-off error becomes the limiting factor of the

method; in fact Euler's method tends to drift further and further from the exact curve (Fig. 4).

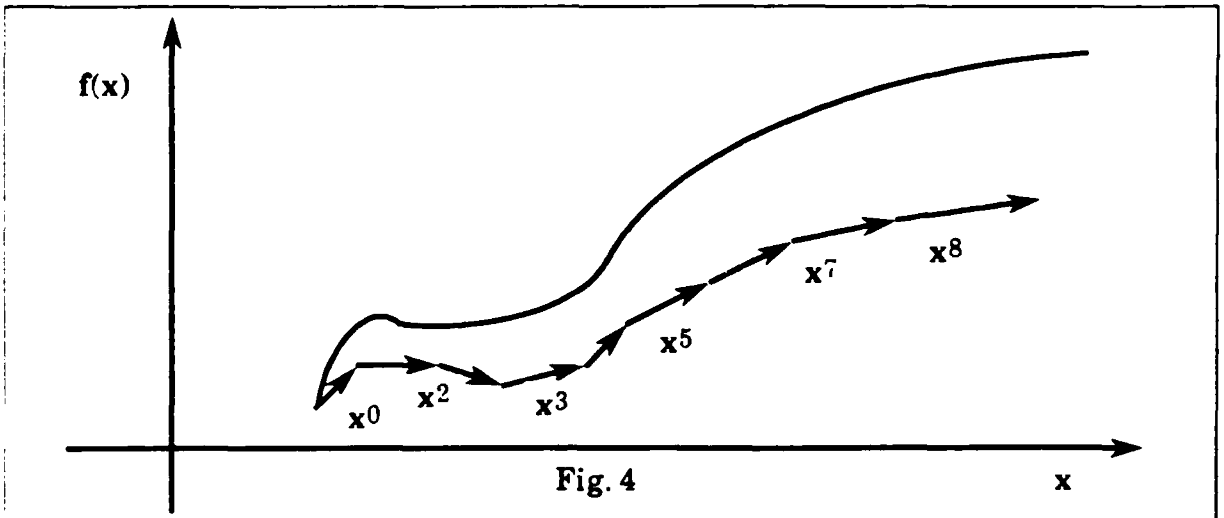


Fig. 4

Euler's method is characterized by a one step computation; in fact it goes from x^k to x^{k+1} without any additional information. Other methods, known as multistep methods, use information from x^k, x^{k-1}, x^{k-2}, x^{k-3}, ... to compute y^{k+1} at x^{k+1}, therefore overcoming much of the drifting of the Euler method ([16]).

A special class of these multistep methods that forms the basis of the software now used for solving differential equations, is Adams methods. they are based on the formula

$$y(x^{k+1}) = y^k + \int_{x^k}^{x^{k+1}} y'(x)dx = \int_{x^k}^{x^{k+1}} f(x, y(x))\,dx$$

One usually models f(x,(y(x))) by a polynomial interpolating y'(x) at various points; then we obtain the Adams formulas from

$$y^{k+1} = y^k + \int_{x^k}^{x^{k+1}} p(x)dx$$

The particular formulas depend on the interpolation points chosen:

a) The Adams-Bashforth method of *order* (≡ degree of the polynomial used) p + 1 uses interpolation of y'(x) at x^k, x^{k-1}, x^{k-2}, x^{k-3}, ... x^{k-p}

b) The Adams-Moulton method of *order* $p+1$ uses interpolation of $y'(x)$ at x^{k+1}, $x^k, x^{k-1}, x^{k-2}, x^{k-3}, \ldots\ x^{k-p+1}$

The first few instances of these formulas are the following:

Adams - Bashforth formulas

order	
1	$y_{k+1} = y_k + hy'_k$ (Euler's method)
2	$y_{k+1} = y_k + (h/2)(3y'_k - y'_{k-1})$
3	$y_{k+1} = y_k + (h/12)(23y'_k - 16y'_{k-1} + {}^*y'_{k-2})$

Adams - Moulton formulas

order	
1	$y_{k+1} = y_k + hy'_{k+1}$
2	$y_{k+1} = y_k + (h/2)(3y'_{k+1} - y'_k)$
3	$y_{k+1} = y_k + (h/12)(5y'_{k+1} - 8y'_k - y'_{k-1})$

5. Getting by Local Extrema

A lot of problems scientists deal with, are formulated as optimization problems. The solutions to these problems can subsequently be visualized as extrema (maxima or minima) of a relevant function. One usually tries, starting from the initial problem, to come up with an equivalent optimization problem that has only one extremum (convexity hypothesis). But this cannot always be the case. In fact, most non-linear problems have equivalent minimization problems with more than one solutions (set of extrema). A subset of the latter consists of the *global extrema*; the rest of this set are points corresponding to local extrema. Application of the above mentioned methods in such a case will definitely give a solution; however it will not, in general, be the one corresponding to the global extremum.

If we can find all of the solutions of a problem, then in many circumstances we can determine the global extremum. The underlying idea is conceptually simple: one finds all of the solutions for the initial problem; assuming that their number is not large, one can sort them and find the one that corresponds to the desired extremum. **The key point is to find all of the solutions.**

In order to this, one has to use a different homotopy than the ones that we mentioned above, called the *all-solutions homotopy* ([2], [9]). It has many different starting

points in $H^{-1}(t^0)$ each of which initiates a different path. Following each of these paths leads to each one of the different solutions.

The difference of the *all-solutions homotopy* from the previously mentioned ones is that now the Hilbert space is based on the complex number field. In other words, E is now of the form $\mathbb{C}^m$ where m is a (positive) natural number. The following theorem provides the necessary framework that supports the *all-solutions homotopy*.

Theorem: Let $F: \mathbb{C}^m \to \mathbb{C}^m$, $F \in C^2$ and suppose that we wish to find all solutions of the system:

$$\mathbf{F(z) = 0}, \text{ where } z \in \mathbb{C}^m$$

We also assume that $\mathbf{F(z) = 0}$ has a finite number of solutions. Suppose now that the homotopy

$$H_i(z,t) = (1-t)((z_i)^{q_i} - 1) + t{*}_i(z), i = 1, 2, 3, 4, \ldots m$$

is regular and that no path can run to infinity for any value of $t \in [t^0, t^1]$. Then by starting from the $Q = \prod_{i=1}^{i=m} q_i$ many solutions of $H^{-1}(t^0)$, and following the corresponding paths, we obtain all solutions of $\mathbf{F(z) = 0}$.†

6. Discussion

To relate the above ideas to Computer Vision we present now an example showing how one can relate the solutions to the Shape from Shading problem through Scale Space. Before we proceed, let us first recall quickly the previous work in this area along with the justification for our choosing this specific example.

---

† The rest of the idea is pretty much expected: after having determined all the specific solutions, one determines which of the pure real solutions obtained results in the optimal value of the corresponding function.

In [2], one can find the following claim: (we quote from page 363):

> "...*Independently and almost simultaneously, Drexler[1977, 1978] and Garcia and Zangwill [1979a, 1980b] developed the all-solutions algorithm*..."

However, the same algorithm was used by Wasserstrom ([9]) in order to find the roots of a 25th degree polynomial, in 1972.

In almost every imaging situation, one has to face the problem of scale; the image under consideration is characterized ([23]) by i) a limited extent (the *outer* scale) and ii) a limited resolution (the *inner* scale). In many applications the *inner* and *outer* scales are set by the imaged object. For example a house does not exist at the scale of the bricks making up its walls, nor at the scale of the city. As Koenderink ([23]) points out:

> *"... if you have no a priori reasons to look for certain features,then you cannot decide on the* **right** *scale, except for certain trivial cases..."*

Hence, if one aims to retain all of the existing structure and still be able to identify objects through (de-)blurring, one must consider the image at different levels of resolution in a continuous way. Another fact that pleads for this approach is the ability of the visual system to "zoom in" on the right range of scale ([21], [22]).

From the above, the necessity for a continuous description of the imaged objects over an interval of resolutions becomes clear, and scale space parameterization ([14]) seems to be the right approach.

Multigrid algorithms have been used in prior work ([24]) in the developing of iterative algorithms for a large number of Computer Vision problems including shape from shading. Although this methodology is characterized by two attractive features such as i) faster convergence to a solution and ii) the computation of mutually consistent visual representations over a range of resolutions, it still requires an *a priori* fixed multigrid structure with a concrete number of levels.

The real challenge ([23]) is to understand and treat the image at all levels of resolution (assuming a closed interval of resolutions, of course) simultaneously and not as an unrelated set of images derived at different scales.

An attempt towards this direction was made by Koenderink ([23]). Motivated by the intuition that the nature of the heat equation ([25]) provides, he embedded the original image in a one parameter family of "derived" images, generated by a diffusion process, and studied the problem by considering the whole family of images. The involved parameter was the scale. He even proceeded further by deriving conditions that had to be satisfied if one wanted to guarantee the accuracy and the stability (in the numerical sense) of the representation.

6.1 Shape from Shading through Scale Space: deriving the relation between the solutions

Consider the shape from shading problem ([19], [26], [27]). This problem can be formulated as an optimization problem as follows:

the solution to the problem is that pair of functions p, q that minimize the value of the following integral:

$$F = \int (\lambda(E - R)^2 + (p_x^2 + p_y^2) + (q_x^2 + q_y^2))\,dxdy \quad (16)$$

where $p = p(x,y)$, $q = q(x,y)$ and (x,y) are the image plane coordinates for any arbitrary point. In the above formulation of the problem we follow the approach taken in [27].

A necessary condition for this integral to have an extremum (in this case a **minimum**) is given by the corresponding Euler equations:

$$(F_1(p,q), F_2(p,q)) = 0$$

where,

$$F_1(p,q) = \theta F/\theta p - \theta(\theta F/\theta p_x)/\theta x - \theta(\theta F/\theta p_y)/\theta y = 0 \quad (17)$$

$$F_2(p,q) = \theta F/\theta q - \theta(\theta F/\theta q_x)/\theta x - \theta(\theta F/\theta q_y)/\theta y = 0 \quad (18)$$

Taking into account equation (16), we can write equations (17) and (18) as:

$$F_1(p,q) = \nabla^2 p + \lambda(\theta R/\theta p)(E - R) = 0 \quad (19)$$

$$F_2(p,q) = \nabla^2 q + \lambda(\theta R/\theta q)(E - R) = 0 \quad (20)$$

Equations (19) and (20) describe our initial non-linear problem. Considering now the following homotopy

$$H(x,\sigma) = F(x,\sigma) = (F_1(p,q,\sigma), F_2(p,q,\sigma)) =$$

$$(\nabla^2 p + \lambda(\theta R/\theta p)(E \otimes G_\sigma - R), \nabla^2 q + \lambda(\theta R/\theta q)(E \otimes G_\sigma - R)) \quad (21)$$

(where $G_\sigma = (1/2\pi\sigma^2)e^{-(x^2+y^2)/2\sigma^2}$ the Gaussian distribution centered at (0,0) with a standard deviation of σ, $\mathbf{x} = (p, q)$, and $\otimes$: the **convolution** operator), in the interval $[0, \sigma_0] = [\sigma_1, \sigma_0]$, we have from (21), that

$$H_1(x,\sigma) = \nabla^2 p + \lambda(\theta R/\theta p)(E \otimes G_\sigma - R) \quad (22)$$

$$H_2(x,\sigma) = \nabla^2 q + \lambda(\theta R/\theta q)(E \otimes G_\sigma - R) \quad (23)$$

Differentiation leads us to:

$$\theta H_1/\theta p = \lambda(\theta^2 R/\theta p^2)(E \otimes G_\sigma - R) - \lambda(\theta R/\theta p)^2 \quad (24)$$

$$\theta H_2/\theta q = \lambda(\theta^2 R/\theta q^2)(E \otimes G_\sigma - R) - \lambda(\theta R/\theta q)^2 \quad (25)$$

$$\theta H_1/\theta q = \lambda(\theta^2 R/\theta q \theta p)(E \otimes G_\sigma - R) - \lambda(\theta R/\theta q)(\theta R/\theta p) \quad (26)$$

$$\theta H_2/\theta p = \lambda(\theta^2 R/\theta p \theta q)(E \otimes G_\sigma - R) - \lambda(\theta R/\theta p)(\theta R/\theta q) \quad (27)$$

$$\theta H_1/\theta\sigma = \lambda(\theta R/\theta p)(-2/\sigma)(E\otimes G_\sigma + I_\sigma) \tag{28}$$

$$\theta H_2/\theta\sigma = \lambda(\theta R/\theta q)(-2/\sigma)(E\otimes G_\sigma + I_\sigma) \tag{29}$$

where

$$I_\sigma = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} - \frac{(x-x')^2 + (y-y')^2}{2\sigma^2}\, e^{-\frac{(x-x')^2 + (y-y')^2}{2\sigma^2}}\, E(x', y')\, dx'\, dy'$$

At this point it is worth mentioning that under the assumption that the partial derivatives $\theta R/\theta p$ and $\theta R/\theta q$ *exist* and are *continuous*, the order of differentiation in $\theta^2 R/\theta p\theta q$ does not matter, i.e. $\theta^2 R/\theta p\theta q = \theta^2 R/\theta q\theta p$. From the above , we can see that the Jacobian is equal to:

$$H' = \begin{bmatrix} \lambda(\theta^2 R/\theta p^2)(E\otimes G_\sigma - R) - \lambda(\theta R/\theta p)^2 & \lambda(\theta^2 R/\theta q\theta p)(E\otimes G_\sigma - R) - \lambda(\theta R/\theta q)(\theta R/\theta p) & \lambda(\theta R/\theta p)(-2/\sigma)(E\otimes G_\sigma + I_\sigma) \\ \lambda(\theta^2 R/\theta p\theta q)(E\otimes G_\sigma - R) - \lambda(\theta R/\theta p)(\theta R/\theta q) & \lambda(\theta^2 R/\theta q^2)(E\otimes G_\sigma - R) - \lambda(\theta R/\theta q)^2 & \lambda(\theta R/\theta q)(-2/\sigma)(E\otimes G_\sigma + I_\sigma) \end{bmatrix}$$

Parameterizing with respect to the length of the arc, and using the Basic Differential Equation approach, as involving mild requirements (see also above discussion) the following system of equations results:

$$dp/ds = -\det \begin{bmatrix} \lambda(\theta^2 R/\theta q\theta p)(E\otimes G_\sigma - R) - \lambda(\theta R/\theta q)(\theta R/\theta p) & \lambda(\theta R/\theta p)(-2/\sigma)(E\otimes G_\sigma + I_\sigma) \\ \lambda(\theta^2 R/\theta q^2)(E\otimes G_\sigma - R) - \lambda(\theta R/\theta q)^2 & \lambda(\theta R/\theta q)(-2/\sigma)(E\otimes G_\sigma + I_\sigma) \end{bmatrix} \tag{30}$$

$$dq/ds = \quad + \det \begin{bmatrix} \lambda(\theta^2R/\theta p^2)(E\otimes G_\sigma - R) - \lambda(\theta R/\theta p)^2 & \lambda(\theta R/\theta p)(-2/\sigma)(E\otimes G_\sigma + I_\sigma) \\ \lambda(\theta^2R/\theta p\theta q)(E\otimes G_\sigma - R) - \lambda(\theta R/\theta p)(\theta R/\theta q) & \lambda(\theta R/\theta q)(-2/\sigma)(E\otimes G_\sigma + I_\sigma) \end{bmatrix} \quad (31)$$

$$d\sigma/ds = -\det \begin{bmatrix} \lambda(\theta^2R/\theta p^2)(E\otimes G_\sigma - R) - \lambda(\theta R/\theta p)^2 & \lambda(\theta^2R/\theta q\theta p)(E\otimes G_\sigma - R) - \lambda(\theta R/\theta q)(\theta R/\theta p) \\ \lambda(\theta^2R/\theta p\theta q)(E\otimes G_\sigma - R) - \lambda(\theta R/\theta p)(\theta R/\theta q) & \lambda(\theta^2R/\theta q^2)(E\otimes G_\sigma - R) - \lambda(\theta R/\theta q)^2 \end{bmatrix} \quad (32)$$

The above derived system of equations, describes the relation of the solutions to the shape from shading problems through scale space.

Recalling equation (21), we observe that the problem to be solved at scale σ_0 is:

$$\nabla^2 p + \lambda(\theta R/\theta p)(E\otimes G_{\sigma_0} - R) = 0 \quad (33a)$$
$$\nabla^2 q + \lambda(\theta R/\theta q)(E\otimes G_{\sigma_0} - R) = 0 \quad (33b)$$

i.e the shape from shading problem for an image derived from the original by blurring with a Gaussian of standard deviation σ_0. Furthermore at scale $\sigma_1 = 0$, the problem to be solved is the one we originally had. Denote by $\mathbf{x}_0$ and $\mathbf{x}_1$ respectively, the solutions to these two problems. Equations (30) - (32) "rule" the path linking the solutions $\mathbf{x}_0$ and $\mathbf{x}_1$. In particular, if one can find a solution to equations (33), one can then track the solution from coarser to finer scales until the solution $\mathbf{x}_1$ is reached. The problem that exists though, is that one cannot guarantee that such an initial solution $\mathbf{x}_0$ can always be found, since one has still to worry about issues of convergence of the method in use.

7. Conclusion and future work

In the above we presented a general framework for the solving of non-linear equations; its power has been demonstrated by an example that showed how one can relate the solutions to the Shape from Shading problem through Scale Space.

We also saw that since there is no guarantee that the initial solution solution x_0 can be always found, relating the solutions to the Shape from Shading problem through Scale Space seems to be the best one can hope for.

However, it may be possible to: i) either derive a homotopy (imbedding) that renders the initial problem at σ_0 always *trivially* solvable, or ii) find a particular value for σ_0 for which the initial problem is always *easily* solvable or iii) use the additional constraint that the above system of equations provide to derive a more robust algorithm for solving the problem at hand.

I am currently being investigating these last three issues; any conclusion at this point would definitely be premature.

It should be clear by now, that the above approach is not restricted to the shape from shading problem only, but can similarly and easily extended to all of the Computer Vision problems that are described by a set of non-linear equations and are subject to scale space formulations.

8. Acknowledgments

I would like to thank my advisor Prof. Christopher Brown for his unconditional support, constructive criticism and for reviewing early drafts of this work, the friend Mike Swain for fruitful discussions and patience, the friend Yiannis Aloimonos for discussions dealing with some technicalities, and the members of the Rochester Vision Group for their stimulating criticism, support and ideas. I would also like to thank the *XEROX Corporation University Grants Program* for providing equipment used in the preparation of this document. This work was supported by a *U.S. Army Engineering Topographic Laboratories Research* grant, number DACA76-85-C-0001.

References:

1) Dieudonne, J., *Foundations of Modern Analysis*, Academic Press 1960.

2) Garcia, C. and W. Zangwill, *Pathways to solutions, fixed points and equilibria*, Prentice Hall, 1981.

3) Den Heijer, C., *The numerical solution of the nonlinear operator equations by imbedding methods*, Mathematical Centre Tract, 1979.

4) Ortega, J. and W. Rheinboldt, *Iterative solutions of nonlinear equations in several variables*, Academic Press, 1970.

5) Klopfenstein, R., "Applications of Differential Equations in General Solving Problems", *Comm ACM*, December 1965.

6) Meyer, G., "On solving nonlinear equations with a one-parameter operator imbedding", *SIAM J.*, December 1968.

7) Ficken, F., "The continuation method for Functional Equations", *Comm. Pure and Applied Math.* 4.

8) Bosarge, W., "Iterative Continuation and the Solution of Nonlinear two-point boundary value problems", *Numerische Mathematik* 17, 1971.

9) Wasserstrom, E., "Numerical solutions by the continuation method", *SIAM Rev.* January 1973.

10) Kubicek, M., "Dependence of Solution of nonlinear systems on a parameter", *ACM Trans. Math. Soft.* 2, March 1976.

11) Dahlquist, G. and A. Bjorck, *Numerical methods*, Prentice Hall, 1974.

12) Backhalov, N., *Numerical Methods*, Mir, Moscow 1977.

13) Rall, L., "Davidenko's method for the solution of nonlinear operator equations", *MRC TR 948*, Un. of Wisconsin, 1968.

14) Witkin, A. "Scale Space Filtering", *Proceedings IJCAI* 1983.

15) Witkin, A., D. Terzopoulos and M. Kass, "Signal Matching through Scale Space", *Proceedings AAAI*, 1986.

16) Rice, J., *Numerical Methods, Software and Analysis*, McGraw-Hill, 1983.

17) Schultz, M., *Differential Geometry*, Schaum's Outline Series.

18) Davidenko, D.F., "On a new method of numerical solution of systems of nonlinear Equations", *Dokl. Akad. Nauk, SSSR*, 88, (1953).

19) Ballard, D. and C. Brown, *Computer Vision*, Prentice-Hall International, 1982.

20) Rigoutsos, I., P. Chou and C. Brown, "Camera Calibration", *Technical Report 186*, University of Rochester, May 1986.

21) Koenderink, J., and A. J. van Doorn, "Visual Deetection of Spatial Contrast: influence of location in the visual field, target extent and illuminance level", *Biol. Cybern.*, 30, 1978.

22) Koenderink, J., and A. J. van Doorn, "Invariant Features of Contrast Detection: an explanation in terms of self-similar detector arrays", *JOSA*, 72, 1982.

23) Koenderink, J., " The Structure of Images", *Biol. Cybern.*, 50, 1984.

24) Terzopoulos, D., "Image Analysis Using Multigrid Relaxation Algorithms", *IEEE PAMI*, 8:2, 1986.

25) Feynman, R., R. Leighton and M.Sands, *The Feynman Lectures on Physics*, Vol.1, Addison-Wesley, 1963.

26) Horn, B.K.P., "Obtaining Shape from Shading Information", in: Winston P. H. (Ed.) *Psychology of Computer Vision*, McGraw Hill, New York, 1975.

27) Ikeuchi, K. and B.K.P.Horn, "Numerical Shape from Shading and Occluding Boundaries", *Artificial Intelligence*, 17, 1981.

END
DATE
3-88
DTIC